Adaptive Mutations that Prevent Crosstalk Enable the Expansion of Paralogous Signaling Protein Families

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1016/j.cell.2012.05.033">http://dx.doi.org/10.1016/j.cell.2012.05.033</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>Elsevier B.V.</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Tue Oct 09 01:54:06 EDT 2018</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/91691">http://hdl.handle.net/1721.1/91691</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>
Adaptive Mutations that Prevent Crosstalk Enable the Expansion of Paralogous Signaling Protein Families

Emily J. Capra,1 Barrett S. Perchuk,1 Jeffrey M. Skerker,3 and Michael T. Laub1,2,*
1Department of Biology
2Howard Hughes Medical Institute
Massachusetts Institute of Technology, Cambridge, MA 02139, USA
3Department of Bioengineering, University of California, Berkeley, Berkeley, CA 94720, USA
*Correspondence: laub@mit.edu
http://dx.doi.org/10.1016/j.cell.2012.05.033

SUMMARY
Orthologous proteins often harbor numerous substitutions, but whether these differences result from neutral or adaptive processes is usually unclear. To tackle this challenge, we examined the divergent evolution of a model bacterial signaling pathway comprising the kinase PhoR and its cognate substrate PhoB. We show that the specificity-determining residues of these proteins are typically under purifying selection but have, in α-proteobacteria, undergone a burst of diversification followed by extended stasis. By reversing mutations that accumulated in an α-proteobacterial PhoR, we demonstrate that these substitutions were adaptive, enabling PhoR to avoid crosstalk with a paralogous pathway that arose specifically in α-proteobacteria. Our findings demonstrate that duplication and the subsequent need to avoid crosstalk strongly influence signaling protein evolution. These results provide a concrete example of how system-wide insulation can be achieved postduplication through a surprisingly limited number of mutations. Our work may help explain the apparent ease with which paralogous protein families expanded in all organisms.

INTRODUCTION
The evolutionary forces and selective pressures that influence protein sequences remain poorly understood at a detailed, molecular level. A comparison of orthologs often reveals tens to hundreds of amino acid differences. How and why do functionally equivalent proteins diverge in different organisms? Many of the accumulated substitutions may be functionally neutral and result from processes such as genetic drift. However, some mutations may have been adaptive and provided a fitness advantage. Identifying these beneficial mutations and pinpointing the advantage that they provide are difficult problems. Comparative sequence analyses, such as measures of codon substitution patterns or dN/dS ratios (Yang and Bielawski, 2000), can help to identify residues that are potentially adaptive, but such approaches are frequently insufficient and difficult to validate. Additionally, elucidating why certain mutations are beneficial requires a genetically manipulatable organism and an ability to probe the effects of individual mutations in vivo.

In many cases where protein evolution has been studied experimentally (reviewed in Dean and Thornton, 2007), the relevant proteins were examined in vitro or in heterologous hosts, and thus outside their native cellular context, possibly eliminating or obscuring important evolutionary constraints. For example, signal transduction proteins are often part of large paralogous families that expand through duplication and divergence. The duplication-divergence process thus runs an inherent risk of introducing crosstalk with existing pathways. A study of SH3 domains from Saccharomyces cerevisiae and humans suggested that the avoidance of crosstalk may represent an important selective pressure in the evolution of paralogous protein families (Zarrinpar et al., 2003). However, a direct demonstration that crosstalk influences the evolution of signaling proteins and, more importantly, an understanding of how this occurs at the amino acid level are lacking.

To tackle these challenges, we examined the evolution of two-component signal transduction proteins in bacteria. These pathways, a primary means of signal transduction in prokaryotes, typically involve a sensor histidine kinase that, upon receipt of an input stimulus, autophosphorylates and then transfers its phosphoryl group to a cognate response regulator, which in turn modulates gene expression (Stock et al., 2000). Most histidine kinases are bifunctional and can, in the absence of an input signal, stimulate the dephosphorylation of their cognate response regulators, effectively acting as phosphatases (Huynh and Stewart, 2011).

Although most bacteria encode between 20 and 200 two-component signaling pathways (Alm et al., 2006), very little crosstalk occurs at the level of phosphotransfer in vivo (Grimes et al., 1998; Laub and Goulian, 2007; Siryaporn and Goulian, 2008; Skerker et al., 2005). Two-component pathways are highly specific, typically with one-to-one relationships between...
cognate kinases and regulators. When not stimulated to auto-phosphorylate, the phosphatase activity of a given histidine kinase can help to eliminate crosstalk and the errant phosphorylation of its cognate regulator (Siryaporn and Goulian, 2008). However, when stimulated as a kinase, molecular recognition is the dominant mechanism for preventing phosphotransfer crosstalk and thereby maintaining the fidelity of distinct signaling pathways. Systematic analyses of phosphotransfer have demonstrated that histidine kinases are endowed with an intrinsic ability to discriminate their in vivo cognate substrate from all other noncognate substrates (Skerker et al., 2005). Analyses of amino acid coevolution in cognate signaling proteins identified the key specificity-determining residues in histidine kinases and response regulators (Bell et al., 2010; Capra et al., 2010; Casino et al., 2009; Skerker et al., 2008; Weigt et al., 2009). Rational mutagenesis of these residues can reprogram the partnering specificity of a histidine kinase or response regulator (Bell et al., 2010; Capra et al., 2010; Skerker et al., 2008).

For a given two-component pathway, there is likely strong purifying selective pressure on its key specificity-determining residues to preserve the kinase-substrate interaction. Even single amino acid changes in specificity residues can drastically change the interaction capabilities and preferences of a histidine kinase or response regulator (Capra et al., 2010). Nevertheless, an inspection of orthologous kinases or response regulators often reveals divergent evolution and variability in the specificity residues of certain subsets of orthologs, raising the question of whether these changes resulted from neutral or adaptive processes. We favored the latter, hypothesizing that specificity residues must change in order to avoid crosstalk between pathways following gene duplication events. Two-component signaling pathways often expand through duplication (Alm et al., 2006), and following such events, bacteria presumably must accumulate mutations that insulate the new pathways from each other and maintain their isolation from other, existing two-component pathways. Here, we provide direct experimental evidence that the avoidance of crosstalk is indeed a major selective force in the evolution of two-component signaling pathways. Through in vitro studies and fitness competition assays, we identify specific substitutions in a model two-component pathway, PhoR-PhoB, that represent an adaptation to the duplication of another two-component signaling system. Similar adaptations likely accompanied each of the duplication and divergence events underlying the massive expansion of two-component signaling protein families in bacteria. Accordingly, global analyses of specificity-determining residues in extant bacterial genomes reveal a pervasive trend toward orthogonality in these signaling proteins.

RESULTS

Vertical Inheritance of PhoR and PhoB

To examine the divergent evolution of two-component signaling pathways, we focused on the PhoR-PhoB signaling pathway (Wanner and Chang, 1987), which is found throughout the bacterial kingdom and helps a wide range of organisms respond to phosphate starvation. To systematically identify orthologs of Escherichia coli phoR and phoB, we used a modified version of reciprocal best hits in BLAST analysis that allows for the identification of putative duplications. Most proteobacteria, except a small number of δ-proteobacteria, were found to encode a single ortholog of each phoR and phoB, suggesting that these genes have rarely been duplicated, particularly in the α-, β-, and γ-proteobacteria (Figure 1A). Additionally, gene trees for phoR and phoB closely matched a species tree (Figure 1B; Figures S1A and S1B available online), indicating that this signaling system has likely been vertically inherited in these clades.

Given that phoR and phoB genes were rarely duplicated during the evolution of proteobacteria, it might be expected that the residues dictating phosphotransfer specificity would be relatively constant in order to preserve the interaction between PhoR and PhoB. We thus examined the six residues in PhoR and seven residues in PhoB previously identified as critical determinants of specificity in two-component signaling proteins (Capra et al., 2010). We extracted these residues, hereafter referred to simply as specificity residues, from 149 PhoR orthologs and 92 PhoB orthologs, and built sequence logos representing the relative frequency of amino acids at each specificity position (Figure 1B; Table S1). The difference in number of PhoR and PhoB orthologs results from the independent identification of kinase and regulator orthologs; most organisms encode both PhoR and PhoB.

The specificity residues of both PhoR and PhoB are generally well conserved (Figures 1B and S1C), although several positions showed substantial variability. We then split the PhoR and PhoB sequences into groups corresponding to the three proteobacterial subdivisions, α, β, and γ. Sequence logos built for each phylogenetic group revealed that differences between subdivisions can account for nearly all of the variability in the combined sequence logos (Figure 1B). For instance, in γ- and β-proteobacteria the first two positions are almost always threonine and valine, whereas in α-proteobacteria these positions are usually alanine and serine or two alanines. Similar observations were made for the specificity residues of PhoB orthologs grouped according to phylogenetic subdivision. Importantly, each PhoR and PhoB sequence logo was built using species that are highly diverged. The strong conservation within each clade thus suggests that specificity residues are usually subject to strong purifying selection. Why, though, have specificity residues diverged between clades?

Identification of Adaptive Mutations that Prevent Crosstalk In Vitro

The clade-specific differences in PhoR and PhoB specificity residues may simply reflect degeneracy in the residues that enable PhoR and PhoB to interact. Alternatively, the differences may have produced functional changes such that a PhoR from one clade is less efficient at interacting with a PhoB from a different clade. To distinguish between these possibilities, we purified PhoR kinases from representative γ- and α-proteobacteria, E. coli and Caulobacter crescentus, and examined their ability to phosphorylate a panel of 11 PhoB orthologs from α, β, and γ-proteobacteria (Figures 1C and S1D). For each PhoB from a γ-proteobacterium, phosphotransfer from the E. coli (γ) PhoR was significantly faster than from C. crescentus (α) PhoR.
Similarly, each PhoB from an α-proteobacterium was preferentially phosphorylated by the α-PhoR. For the two chosen β-PhoR orthologs, we observed more rapid phosphorylation by the γ-PhoR than the α-PhoR, consistent with the specificity residues of the β-PhoR and β-PhoB orthologs being more similar to those found in γ-proteobacteria than those in α-proteobacteria. We conclude that within each proteobacterial subdivision, the phosphotransfer specificity of PhoR and PhoB orthologs is relatively static. However, substitutions in the specificity residues of α-PhoR and α-PhoB orthologs have led to significant differences in phosphotransfer specificity between clades.

The changes in PhoR-PhoB specificity residues, and consequent alteration of interaction specificity could have resulted from neutral drift. However, the strong conservation of specificity residues within each clade, which includes species that are widely divergent, suggests that such drift is extremely rare. Instead, the alternative PhoR-PhoB specificity residues in α-proteobacteria may be adaptive and provide an important selective advantage. We hypothesized that the substitutions in α-PhoR and α-PhoB specificity residues prevent unwanted crosstalk with another pathway that is specific to the α-proteobacteria, i.e., negative selection led to changes in α-PhoR and α-PhoB. This model predicts that PhoR orthologs from γ-proteobacteria may phosphorylate response regulators found exclusively in α-proteobacteria, which the α-PhoR orthologs have adapted to avoid phosphorylating.

To test this possibility, we performed comprehensive phosphotransfer profiling of E. coli (γ) and C. crescentus (α) PhoR. Both PhoR constructs were autophosphorylated in vitro and then examined, in parallel, for phosphotransfer to the 44 response regulators encoded by C. crescentus (Figure 2). Both PhoR constructs phosphorylated the C. crescentus PhoB, consistent with their orthologous relationship; although as noted above, phosphotransfer from the α-PhoR is more robust. Interestingly, the γ-PhoR showed significant phosphotransfer to NtrX, whereas the α-PhoR construct did not. Notably, most

Figure 1. Phosphotransfer Specificity of PhoR Is Different in α- and γ-Proteobacteria
(A) Percentage of genomes harboring one or two Pho pathways.
(B) Neighbor-joining tree for a subset of PhoR orthologs and sequence logos for specificity residues in PhoR and PhoB orthologs. Logos are shown for orthologs in each subdivision and as a combined set. Bootstrap values are out of 1,000. A neighbor-joining tree for PhoB orthologs and a species tree for the species used are in Figures S1A and S1B.
(C) Time courses of phosphotransfer from C. crescentus PhoR (α) and E. coli PhoR (γ) to each of 11 PhoB orthologs from representative α-, γ-, and β-proteobacteria as noted above each graph. Band intensities for each PhoB were normalized in each experiment to the initial amount of autophosphorylated kinase. Values for PhoB phosphorylation can be greater than one as ATP was in excess in the reaction, allowing for reautophosphorylation of PhoR and subsequent transfer to PhoB. For original gel images, see Figure S1D.
α-proteobacteria encode two paralogous Ntr systems, NtrB-NtrC and NtrX-NtrY, while the γ-proteobacteria typically encode only one, NtrB-NtrC (Figure 3A; Table S1). The two α-Ntr systems, which likely arose through duplication and divergence, do not crosstalk with each other in vitro (Figure 3B) and, consistently, have different specificity residues (Figure 3C). Collectively,

Figure 2. Substituting γ-like Specificity Residues into α-PhoR Increases Phosphorylation of NtrX
(A) Phosphotransfer profiling of E. coli PhoR, C. crescentus PhoR, and C. crescentus PhoR mutants containing γ-proteobacterial specificity residues. Each autophosphorylated kinase, indicated on the right, was incubated with each of 44 C. crescentus response regulators, indicated across the top, for 15 min.
(B) Quantification of profiles in (A). Band intensities for each response regulator were normalized to the level of autophosphorylated kinase and then plotted relative to PhoB.
Also see Figures S2D and S2E.
we focused on PhoB and NtrX as time courses of phosphotransfer from PhoR kinases phosphorylated several substrates (see Figure 2), although see Figures S2A–S2C. Error bars represent standard deviations, n = 3.

Strikingly, each mutant led to a significant increase in NtrX phosphorylation (Figure 2). To test whether these mutations also prevent crosstalk in vivo, we engineered the chromosomal copy of phoR (Figures S2F and S2G). The most significant crosstalk to NtrX occurred for PhoR mutants with a valine substituted for serine at specificity position 2. Importantly, substantial crosstalk was not observed when this serine was substituted with other residues including leucine, aspartate, glutamate, and threonine. Only valine, corresponding to that found in γ-proteobacterial PhoR orthologs, produced significant crosstalk (Figures S2F and S2G).

Taken together, our in vitro studies support the notion that alanine, serine, and phenylalanine at specificity positions 1, 2, and 4 represent adaptive mutations that prevent crosstalk to NtrX. To test this hypothesis, we created a series of C. crescentus PhoR mutants in which specificity residues were replaced with the corresponding residues from γ-proteobacterial PhoR. We made each single mutant, three double mutants, and the triple mutant. Each mutant kinase was then profiled against the complete set of C. crescentus response regulators to examine what effect, if any, these residues have on phosphotransfer specificity. Strikingly, each mutant led to a significant increase in NtrX phosphorylation (Figure 2).

We also examined detailed time courses of phosphotransfer from each mutant PhoR, as well as the wild-type kinases, to the C. crescentus regulators PhoB and NtrX. Each mutant kinase exhibited an increase in crosstalk with NtrX compared to the wild-type C. crescentus PhoR, but retained the ability to phosphorylate C. crescentus PhoB at rates comparable to the wild-type PhoR (Figures 3D and S2A–S2C). Although some mutant PhoR kinases phosphorylated several substrates (see Figure 2), we focused on PhoB and NtrX as time courses of phosphotransfer indicated these as the two preferred targets of mutant PhoR constructs (Figures S2D and S2E).

Avoidance of Crosstalk Is a Significant Selective Pressure
To test whether these mutations also prevent crosstalk in vivo, we engineered the chromosomal copy of phoR in the γ-proteobacterium C. crescentus to produce a mutant PhoR in which specificity positions 1 and 2 are threonine and valine, respectively, as they are in γ-proteobacteria; hereafter, this mutant strain is referred to as PhoR(TV). Based on our in vitro experiments, we expected that crosstalk from PhoR(TV) to NtrX would be induced during growth in phosphate-limited media (Figure 4A). During growth in such conditions, wild-type PhoR is stimulated to autophosphorylate and phosphotransfer to PhoB, which then activates genes involved in responding to
phosphate limitation. Thus, any effects of increased crosstalk to NtrX by the PhoR(TV) kinase should be manifest specifically during growth in phosphate-limited media.

We grew cells to midlogarithmic phase in phosphate-limited media and measured the rate of growth by monitoring the accumulation of optical density at 600 nm. In minimal media containing either 50 μM phosphate or 5 μM phosphate, the PhoR(TV) mutant grew significantly more slowly than wild-type, with a doubling time ≈30% longer than wild-type in each case (Figures 4B, S3A, and S3B). This growth defect was almost as severe as that observed for a ΔphoR strain which cannot mount a proper transcriptional response to phosphate limitation. To assess whether crosstalk from PhoR(TV) to NtrX contributed to the slow growth phenotype observed, we deleted ntrX in the PhoR(TV) strain. Indeed, the deletion of ntrX significantly reduced the growth deficiency of the PhoR(TV) mutant (Figures 4B, S3A, and S3B) suggesting that crosstalk with NtrX contributes significantly to the slow growth phenotype of a PhoR(TV) strain. The suppression observed was not a nonspecific acceleration of growth as the ntrX deletion alone had no effect on growth in phosphate-limited medium.

In phosphate-replete medium, the PhoR(TV) mutant strains grew at a rate nearly identical to the wild-type (Figure 4B), indicating that, as expected, crosstalk to NtrX requires PhoR to be activated as a kinase. The ntrX deletion and PhoR(TV)/ΔntrX strains grew more slowly in phosphate-replete medium, as the NtrY-NtrX pathway is likely necessary for responding to a signal or metabolite produced in M2G medium.

To corroborate our growth rate measurements, we performed competitive fitness assays in which each mutant strain was mixed with the wild-type at a ratio of 1:1 and grown in the same flask for 104 hr, or approximately 40 wild-type generations.
similarly in the PhoR(TV) and significantly affected in the PhoR(TV) strain during growth in the growth defect of the PhoR(TV) mutant. However, NtrX-inappropriate phosphorylation of NtrX could also contribute to a model in which phosphorylation crosstalk from PhoR(TV) to the kinase. Collectively, these data further support a model in which into PhoR likely occurs only in conditions in which PhoR is a positive disadvantage of introducing ancestral specificity residues into PhoR likely occurs only in conditions in which PhoR is a kinase. Collectively, these data further support a model in which the $\alpha$-specific substitutions in PhoR specificity residues (T $\rightarrow$ A and V $\rightarrow$ S at specificity positions 1 and 2) are selectively advantageous because they help prevent phosphotransfer crosstalk to NtrX, and perhaps other response regulators.

The growth and competitive fitness defects of PhoR(TV) in phosphate-limited media were comparable to that seen for $\Delta$phoR. This similarity suggested that the detrimental effect of crosstalk in the PhoR(TV) strain stems from an inability to phosphorylate PhoB and activate PhoB-dependent genes in phosphate-limited conditions. To test this hypothesis directly, we examined global gene expression patterns in the PhoR(TV) and $\Delta$phoR strains during growth in phosphate-limited conditions. These expression profiles exhibited strong similarity with a Pearson correlation coefficient of $-0.9$ (Table S2), supporting a model in which phosphorylation crosstalk from PhoR(TV) to NtrX comes at the expense of phosphorylating PhoB. The inappropriate phosphorylation of NtrX could also contribute to the growth defect of the PhoR(TV) mutant. However, NtrX-dependent genes (see Experimental Procedures) were not significantly affected in the PhoR(TV) strain during growth in phosphate-limited conditions; NtrX-dependent genes behaved similarly in the PhoR(TV) and $\Delta$phoR strains in phosphate-limited conditions (Table S2). This may result from NtrY, the cognate kinase for NtrX, functioning as a phosphatase to prevent the accumulation of phosphorylated NtrX in phosphate-limited media. Consistent with this notion, ntrX and ntrY are not required for growth in phosphate-limited media, suggesting that in this condition NtrY is likely in a phosphatase state.

Importantly, and in contrast to NtrY, PhoR functions as a kinase, not a phosphatase, in phosphate-limited media. Thus, our results indicate that the $\alpha$-specific substitutions in PhoR specificity residues (T $\rightarrow$ A and V $\rightarrow$ S) impact fitness by affecting crosstalk at the level of phosphotransfer. Consistently, in phosphate-replete media, when PhoR is primarily active as a phosphatase, these substitutions had little to no effect on competitive fitness (Figures 4B, 4C, S3C, and S3D). To further confirm that these substitutions do not significantly impact the phosphatase activity of PhoR, we examined global patterns of gene expression in the PhoR(TV) mutant grown in a phosphate-replete medium. Under these conditions, PhoR likely acts as a phosphatase to eliminate any errant phosphorylation of PhoB. Accordingly, the expression levels of known PhoB-dependent genes, such as $\text{pstC}$, $\text{pstA}$, and $\text{pstB}$, were modestly elevated in a $\Delta$phoR strain grown in phosphate-replete medium (Figure 4D). By contrast, these genes were not affected, or were slightly downregulated, in the PhoR(TV) strain grown in the same phosphate-replete conditions, indicating that PhoR(TV) retains phosphatase activity in vivo. Collectively, our data demonstrate that the growth and fitness defect of the PhoR(TV) mutant stems from inappropriate phosphotransfer to NtrX, and perhaps other noncognate substrates.

Different Adaptive Mutations Prevent Crosstalk in Other Proteobacterial Clades

Our results suggest that $\alpha$-proteobacteria have accumulated substitutions in PhoR that prevent unwanted crosstalk with the noncognate substrate NtrX. There could, however, be other ways to avoid crosstalk between these systems in other clades. Like the $\alpha$-proteobacteria, most $\beta$-proteobacteria encode NtrY-NtrX orthologs (Figure 3C). However, the $\gamma$-PhoR orthologs have specificity residues at positions 1 and 2, similar to those found in $\gamma$-PhoR orthologs. This observation suggests that either the $\beta$-proteobacteria can tolerate crosstalk between PhoR and NtrX, or other mutations have emerged to prevent PhoR from phosphorylating NtrX. We favored the latter possibility as a comparison of sequence logos for the NtrX orthologs from $\alpha$- and $\beta$-proteobacteria revealed differences at two critical positions (Figure 3C). Whereas most $\alpha$-NtrX orthologs have aspartate, glycine, and lysine at specificity positions 2, 5, and 7, respectively, the $\beta$-NtrX orthologs typically have glycine, glutamate, and alanine at these same three respective positions. We speculated that the different specificity residues in a given $\beta$-NtrX may eliminate crosstalk from a $\beta$-PhoR; that is, $\beta$-proteobacteria may have evolved to avoid crosstalk by accumulating substitutions in NtrX rather than PhoR and PhoB.

To test this hypothesis, we asked whether introducing the $\beta$-NtrX specificity residues into an $\alpha$-NtrX would eliminate crosstalk from $\alpha$-PhoR(TV) which, as shown above, phosphotransfers to $\alpha$-NtrX in vitro and in vivo. Indeed, whereas C. crescentus NtrX was robustly phosphorylated by PhoR(TV), a mutant NtrX harboring the $\beta$-like substitutions D13G, G20E, F107I, and K108A was not detectably phosphorylated (Figure 4E). This mutant NtrX was not simply unfolded or unphosphorylatable as it was still robustly phosphorylated by $\alpha$-NtrY. Hence, the substitutions introduced specifically eliminated crosstalk from PhoR(TV), while still allowing for interaction with the cognate kinase NtrY. Taken together, these results suggest that in $\beta$-proteobacteria, substitutions in NtrX alleviated crosstalk with PhoR while in $\alpha$-proteobacteria substitutions in PhoR prevented crosstalk with NtrX. Although the substitutions are different, the net result in both cases was an insulation of the Ntr and Pho systems.

Global Optimization of Signaling Fidelity

Our results with the Pho and Ntr signaling pathways indicate that the avoidance of crosstalk following gene duplication is a major selective pressure that drives the accumulation of adaptive substitutions in the specificity-determining residues of two-component signaling proteins. More generally, this model
predicts that the specificity residues of two-component signaling proteins in extant organisms should be sufficiently different from, or orthogonal to, one another to prevent crosstalk. To test this prediction, we extracted the six major specificity residues from each of the 22 canonical histidine kinases encoded in the E. coli K12 genome (Figure 5A). Pairwise comparisons indicated that kinases typically had no more than three identities with every other kinase at these six specificity sites, often with nonconservative differences at the remaining sites. One notable exception is NarX and NarQ, which contain two identities and four conservative differences. However, these kinases, which likely arose through gene duplication, each phosphorylate the response regulators NarL and NarP in vitro and likely in vivo, and hence represent a case of physiologically beneficial cross-regulation (Noriega et al., 2010). Aside from these two kinases, there is a general pattern of orthogonality between specificity residues in the system-wide set of E. coli histidine kinases. This orthogonality is further reflected by a lack of information in a sequence logo built from the specificity residues of the 22 E. coli histidine kinases (Figure 5B), particularly in comparison to the sequence logos built from orthologous histidine kinases (Figures 1B and 3C). A similar pattern of orthogonality was evident in the specificity residues of the 20 canonical histidine kinases in C. crescentus, as well as the specificity residues of the response regulators from both E. coli and C. crescentus (Figure S4). These observations, in combination with our detailed investigation of the Ntr and Pho proteins across phylogenies, suggest that the avoidance of crosstalk is a pervasive and significant selective pressure driving the system-wide insulation of two-component signaling pathways, and consequently, that in extant organisms, two-component systems are largely insulated from one another (Figure 5C).

**DISCUSSION**

Signaling protein families, in both prokaryotes and eukaryotes, frequently expand through gene duplication (Ohno, 1970; Pires-daSilva and Sommer, 2003). The retention of the duplicated genes often requires mutations that insulate them from one another, allowing each to transmit signals without inducing crosstalk. This divergence process may be additionally constrained by a need to avoid crosstalk with other, existing members of the same protein family. For two-component pathways, the duplication-divergence process can be conceptually framed by considering the sequence space defined by the specificity-determining residues of histidine kinases (Figures 5 and 6). For each kinase, these residues dictate the substrates it can phosphorylate, with different kinases recognizing largely nonoverlapping, or orthogonal, sets of substrates. Gene duplication leads initially to a complete overlap and requires that one or both of the duplicates accumulate changes in its specificity residues, thus separating them in sequence space (Figure 6).

The mutational path taken by a given kinase may cause it to infringe on the sequence space occupied by another kinase, as was likely the case with the Ntr and Pho systems in z-proteobacteria. Such overlap then necessitates additional mutations to achieve a system-wide optimization of specificity. Our results indicate that such optimization and the avoidance of crosstalk are important selective pressures influencing two-component pathways.
systems and they can drive the divergent evolution of orthologous proteins.

How did the NtrY-NtrX pathway arise if crosstalk is detrimental? Although we cannot infer the order of events with complete certainty, a plausible scenario is that the NtrY-NtrX pathway arose during growth in phosphate-replete conditions where it provided a selective advantage, as suggested by the slow growth of a ΔntrX strain in these conditions. Subsequent growth in phosphate-limited conditions would then select for strains that had accumulated mutations eliminating crosstalk between PhoR and NtrX. We showed that such insulation can occur with only one or two point mutations, supporting the plausibility of this scenario in an ancestral α-proteobacterium. Interestingly, the β-proteobacteria likely followed a different mutational path to avoiding crosstalk, accumulating substitutions in NtrX rather than PhoR. The difference between the two clades of proteobacteria may reflect the inherent stochasticity of mutations and selection. Alternatively, the growth conditions or genomic context of the ancestral organisms in which gene duplication occurred may have deterministically influenced selection.

In sum, we propose that the evolution of two-component signaling genes is characterized by long periods of stasis with specificity-determining residues subject to strong purifying selection to ensure robust phosphotransfer from kinase to regulator. Gene duplication, or lateral transfer events, can disrupt this stasis, requiring a global reoptimization of existing signaling proteins to accommodate the new pathway. The specificity residues are thus likely subject to bursts of diversifying selection; however, these residues would not necessarily exhibit commonly used signatures of diversifying selection such as large dN/dS values. Instead, our work emphasizes that a molecular-level understanding of protein evolution and the identification of adaptive mutations ultimately demands an integration of sequence analysis with focused biochemical and genetic characterizations.

Our approach and findings are relevant beyond two-component signaling as paralogous signaling protein families are found throughout biology. In fact, most organisms use a remarkably small number of types of signaling protein to carry out their diverse information-processing tasks. In all cases, duplication and divergence is a primary means by which new pathways are created and, consequently, issues of specificity and the fidelity of information transfer are critical. While eukaryotes sometimes rely on tissue-specific expression of paralogous genes or spatial mechanisms like scaffolds to enforce specificity, many common signaling proteins and domains, such as PDZ, SH3, SH2, and bZIP proteins (Hou et al., 2009; Liu et al., 2011; Newman and Keating, 2003; Stiffler et al., 2007; Tonikian et al., 2008; Zarrinpar et al., 2003), rely on molecular recognition and a relatively small set of specificity-determining residues. Hence, our observation that pathway insulation in bacteria can be achieved with a limited number of mutations may help to explain how organisms in all domains of life have exploited gene duplication to expand and diversify their signaling repertoires.

EXPERIMENTAL PROCEDURES

Identification of Orthologs and Construction of Gene Trees
A modified version of reciprocal best blast hits was used to identify orthologous proteins. For E. coli PhoR, the DHp domain was used as a query in BLAST searches against fully sequenced bacterial genomes in GenBank (September 2009). The top ten hits from each genome were then subjected to reciprocal BLAST searches against the E. coli MG1655 genome. If only the top hit identified E. coli PhoR as the best match, it was called as a PhoR ortholog. If multiple hits identified E. coli PhoR as the best match, the top hit was called as an ortholog and additional hits were evaluated as follows. If an additional hit had an E-value within 103 of the top hit and was closer to the top hit than to the fifth hit (which generally had an E-value reflecting the overall paralogous relationship of histidine kinases), we also called it as an ortholog of E. coli PhoR and examined the next hit similarly. For genomes with more than one hit called as an ortholog, duplications were inferred and each hit deemed a member of the PhoR ortholog. A similar procedure was followed to identify orthologs of PhoB, NtrB, NtrC, NtrX, and NtrY using as query sequences the E. coli PhoB receiver domain, the C. crescentus NtrB and NtrY DHp domains, or the C. crescentus NtrC and NtrX receiver domains.
Orthologous sequences were aligned using ClustalX (Chenna et al., 2003). Sequence logos for the specificity residues, extracted from the aligned sequences, were built using WebLogo (Crooks et al., 2004). To help correct for phylogenetic biases in genome sequencing efforts, sequences were filtered to ensure that no two sequences were more than 95% identical.

PhoR DHP domains and PhoB receiver domains were extracted using HMMER with models for a His_KA domain and no REC domain. Reported bootstrap values are out of 1,000.

In Figure 1C, each PhoB construct contained only the receiver domain. For time courses of phosphotransfer profiles against all regulators except for CC1741 (C. crescentus), CC1743 (C. crescentus), ntrX, and cenR, gel images were run in parallel and exposed to the same phosphorscreen. Gel images were analyzed using ImageJ. Competition experiments were done once with wild-type expressing cfp and mutant expressing yfp, repeated with fluorescent proteins swapped, and results averaged.

Microarray Analysis
Cultures were grown to mid-log phase in M2G and either RNA was harvested or cells were washed, resuspended in M5G, and grown for 11 hr in phosphate-limited conditions before RNA was harvested. RNA was extracted, labeled, and hybridized to custom-designed 8x15K Agilent expression arrays as described previously (Gora et al., 2010). NtrX-dependent genes were defined as those genes exhibiting at least a 4-fold decrease in expression in the ntrX strain compared to wild-type C. crescentus in M2G. Complete array data are provided in Table S2 and deposited in GEO.

SUPPLEMENTAL INFORMATION
Supplemental Information includes four figures and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2012.05.033.

ACKNOWLEDGMENTS
We thank O. Ashenberg for help with bioinformatics, Y.E. Chen for help with strain construction, and A. Podgornaia, A. Keating, O. Ashenberg, and K. Foster for helpful comments on the manuscript. Sequence analyses were performed on a computer cluster supported by NSF grant 0821391. M.T.L. is an Early Career Scientist of the Howard Hughes Medical Institute. This work was supported by an NSF graduate fellowship to E.J.C. and an NSF CAREER award to M.T.L.

REFERENCES


